

43722nm3.doc

11/10/01

A Method and System for Voice Browsing Web Sites

REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority from U.S. Provisional Application Serial No. 60/243,244 entitled: "A method and system for voice browsing web sites" and filed on October 25, 2000.

FIELD OF THE INVENTION

[0002] The present invention relates to wireless, voice-activated access to information residing on the Internet.

BACKGROUND OF INVENTION

[0003] The advent of the Internet has enabled more rapid publication of a wealth of information to wider audiences than ever before, at significantly lower costs. Over the last ten years tremendous efforts have been made to publish information in HTML, which is easily accessible to anyone with a computer, a web browser and an Internet connection. More recently, the introduction of HDML and the subsequent introduction of WML have enabled mobile users to access published information using hand-held wireless devices.

[0004] Wireless browsers have increased access to Internet-published information for a small segment of the population. WAP (Wireless Application Protocol) enabled devices enable users to access web based information instantly via mobile telephones, pagers, two-way radios, smart phones and communicators. Handheld PDAs (Personal Digital Assistants) also enable users to access web based information, usually by first downloading an application file from a relevant web site.

[0005] For the large remainder of the population who do not have access to a WAP enabled device or PDA, the introduction of Interactive Voice Response Units (IVR's) connected to the Internet has enabled access to web based information from any telephone.

SUMMARY OF THE INVENTION

[0006] Although an IVR may be capable of accessing information that resides on the Internet, there is a lack of methodology to automatically construct audio content from textual format residing on the Internet.

[0007] There is thus provided in accordance with a preferred embodiment of the present invention a method for automatic conversion of text to speech including automatically analyzing a text to define at least one vocabulary domain and carrying out a text-to-speech conversion by employing said at least one vocabulary domain.

[0008] There is also provided in accordance with a preferred embodiment of the present invention a system for automatic conversion of text to speech, which includes an automatic text analyzer and vocabulary domain definer, automatically analyzing a text to define at least one vocabulary domain and a text-to-speech converter, carrying out a text-to-speech conversion by employing said at least one vocabulary domain.

[0009] Further in accordance with a preferred embodiment of the present invention the step of automatically analyzing includes utilizing a closeness metric for defining said at least one vocabulary domain. Preferably, the closeness metric is a content-based metric.

[0010] Still further in accordance with a preferred embodiment of the present invention the method also includes transmitting speech resulting from said text-to-speech conversion over a telephone link.

[0011] Additionally in accordance with a preferred embodiment of the present invention the step of automatically analyzing text comprises analyzing a text published on a web site.

[0012] Additionally or alternatively, the step of automatically analyzing text comprises generating speech recognition grammar.

[0013] Further in accordance with a preferred embodiment of the present invention the step of automatically analyzing text comprises comparing a newly defined vocabulary domain with at least one previously defined vocabulary domain.

[0014] Still further in accordance with a preferred embodiment of the present invention the method operates to convert at least one of HDML, HTML and WML format texts to at least one of VXML, and VoiceXML.

[0015] Additionally in accordance with a preferred embodiment of the present invention the step of carrying out a text-to-speech conversion employs multiple text-to-speech converters.

[0016] Further in accordance with a preferred embodiment of the present invention the system for automatic conversion of text to speech includes multiple text-to-speech converters, at least two of which correspond to at least two different vocabulary domains.

[0017] There is further provided in accordance with a preferred embodiment of the present invention a method for automatic conversion of text to speech including the steps of carrying out a text-to-speech conversion by employ multiple text-to-speech converters, at least two of which correspond to at least two different vocabulary domain and carrying out a text-to-speech conversion by employing said at least one vocabulary domain.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] The present invention will be more fully understood and appreciated from the following detailed description, taken in conjunction with the drawings, in which:

[0019] Fig. 1 is a simplified illustration of a method and system for preparation of an existing textual Internet page, for future audio publication;

[0020] Fig. 2 is a simplified illustration of a method and system for audio publication of textual information on a web site; and

[0021] Fig. 3 is a simplified illustration of the function and operation of one embodiment of a text-to-speech server forming part of the embodiment of Fig. 2.

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

[0022] The present invention provides a system and methodology for converting and delivering textual information, typically including menus and content, such as Wireless Application Protocol (WAP) enabled information.

[0023] In a typical scenario, in accordance with the present invention, a Service Provider may wish to voice-enable textual information, such as local weather or news,

for access thereto over the telephone. The process of voice-enabling an existing text based web site preferably comprises the following three steps:

[0024] First, the Service Provider specifies the location of the textual information. The Service Provider may connect via a standard web browser to the system of the present invention. The Service Provider may then fill out a form specifying a relevant URL such as an HDML/WML/HTML web site in order to receive textual information such as a weather report.

[0025] Next, the Service Provider may receive an acknowledgment page that may contain, among other information, the Service Provider's uniquely assigned Direct Inward Dial (DID) number.

[0026] Finally, a subscriber may place a telephone call to the assigned DID number in order to access the system of the present invention. The textual information provided by the Service Provider may then be retrieved and broadcast to the subscriber over the telephone.

[0027] Reference is now made to Fig. 1, which illustrates a system and methodology for preparation of an existing textual Internet page for future broadcast. A Service Provider 100 may connect to a TTS HTTP server 110 by utilizing a web browser and may retrieve a form. The Service Provider 100 may fill out the form specifying the location of the textual information, typically the URL of an HDML/WML/HTML web site located on a Service Provider HTTP Server 120. Optionally, the Service Provider 100 may also specify audio content that may be placed in an Audio Database 130. Should the Service Provider 100 submit the form to the HTTP Server 110, the TTS HTTP server 110 may connect to a DID Database 140 to retrieve a DID number and may assign it to the Service Provider 100. The TTS HTTP server 110 may return an acknowledgement page to the Service Provider 100 that may contain, among other information, the DID number assigned to the Service Provider 100.

[0028] The TTS HTTP server 110 may forward the location of the textual information, typically the URL, to an Analyzer/Vocabulary Domain Definer 150 to be analyzed. The Analyzer/Vocabulary Domain Definer 150 may connect to the Service Provider HTTP Server 120 and request the URL. The Analyzer/Vocabulary Domain Definer 150 may then span the various HDML/WML/HTML pages found on the

Service Provider HTTP Server 120, following hyperlinks and collecting the vocabulary of the textual information published thereon.

[0029] The Analyzer/Vocabulary Domain Definer 150 may further analyze the assembled vocabulary to determine a lexicon and vocabulary domains represented thereby. A web site may contain text that can be grouped into different limited vocabulary domains, in which each limited domain contains a cluster of textual information including at least partially similar vocabularies. For example, the Analyzer/Vocabulary Domain Definer 150 may group sentences that share one or more selected words into the same limited vocabulary domain. Thus, for example, all published textual information regarding "weather" may be placed into a single limited vocabulary domain. Similarly, all queries such as forms regarding "city-state information" or "customer information" may define different limited vocabulary domains.

[0030] Once the textual information has been clustered into its respective limited vocabulary domains, similar textual information received in the future may be mapped to respective clusters within appropriate vocabulary domains.

[0031] The Analyzer/Vocabulary Domain Definer 150 may compare the vocabulary domains required to represent the textual information of the web site with existing recorded audio, stored in the Audio Database 130. Should the Analyzer/Vocabulary Domain Definer 150 determine the need to record new audio files, the Analyzer/Vocabulary Domain Definer 150 may send a request to a Recording Studio 160 with the sentences or words to be recorded. The Recording Studio 160 provides the Audio Database 130 with the sentences and/or words recorded. The complete set of formatting configuration information necessary to format the textual web site for audio publication may be stored for later retrieval in a User Database 170. At the time of such retrieval, as described in more detail in Fig. 2, an IVR 180 may access the textual information on the Service Provider HTTP Server 120 and may convert the textual information to audio on the fly, by utilizing the User Database 170.

[0032] Optionally, if the Service Provider 100 specifies audio content, an Audio Distributor 190 may distribute specified audio files to one or more IVRs 180. In this situation each IVR 180 may access specified audio files locally, such as from the IVR's hard drive.

[0033] Reference is now made to Fig. 2, which illustrates a method and system employed during retrieval to format a textual web site for audio publication. A Subscriber 200, typically employing a telephone, communicates with an IVR 180. The IVR 180 may be employed to access textual information published on the Service Provider HTTP Server 120. This may be accomplished either by the Subscriber 200 explicitly specifying the textual information. Alternatively, the IVR 180 may detect the preferences of the Subscriber 200 either through Dialed Number Identification Service (DNIS) or Automatic Number Identification (ANI).

[0034] Next, the IVR 180 may request to retrieve the textual information from a Vocabulary Domain Based Text-to-Speech Converter 210. The Vocabulary Domain Based Text-to-Speech Converter 210 may connect to the Service Provider HTTP Server 120 and may request the textual information. The Service Provider HTTP Server 120 may transmit the textual information, such as HDML/WML/HTML information to the Vocabulary Domain Based Text-to-Speech Converter 210. The Vocabulary Domain Based Text-to-Speech Converter 210 may also retrieve the previously defined formatting configuration information from the User Database 170, and employ the formatting configuration information to convert the textual information retrieved from Service Provider HTTP Server 120 into a mark up language that the IVR 180 may process, such as VoiceXML®.

[0035] During the process of conversion, the Vocabulary Domain Based Text-to-Speech Converter 210 may further utilize the formatting configuration information to insure that the IVR 180 will make efficient use of a Text to Speech Server (TTS) 220. This may be accomplished through mapping the text to clusters, previously defined in a preparatory stage described hereinabove with reference to Fig. 1. Should the Vocabulary Domain Based Text-to-Speech Converter 210 fail to map or parse the textual information, for example should the textual information on the Service Provider HTTP Server 120 have changed dramatically from a previous communication with the web site, the Vocabulary Domain Based Text-to-Speech Converter 210 preferably notifies the Analyzer/Vocabulary Domain Definer 150 (Fig. 1). The Analyzer/Vocabulary Domain Definer 150, upon receiving a notification of changed textual information on the web site, may analyze the web site as previously described in the preparatory phase described hereinabove with reference to Fig. and transfer the new

textual information to the Audio Database 130 and/or to the Recording Studio 160. Additionally the Analyzer/Vocabulary Domain Definer 150 may send an email notification to the Service Provider 100 (Fig. 1).

[0036] While providing service to the Subscriber 200, the IVR 180 may remain in contact with a License Manager 230 throughout. The License Manager 230 is responsible for ensuring that subscribers are billed in accordance with usage. The License Manager 230 may retrieve subscriber configuration information from the User Database 170 and monitor subscriber usage. This methodology enables the IVR 180 to interrupt the Subscriber 200, should the License Manager 230 determine that subscriber 200 has exceeded any previously specified limits set by the Service Provider 100 (Fig. 1), such as pre-paid calling time limits.

[0037] Optionally, the Service Provider 100 (Fig. 1) may configure the textual information residing on the Service Provider HTTP Server 120 to incorporate a proprietary API (not shown) that may enable the Vocabulary Domain Based Text-to-Speech Converter 210 to fully utilize the mark-up language. For instance, the Service Provider 100 may possess pre-recorded audio that resides on a Proprietary HTTP Server 125, that describes the current news in Pakistan. When the Subscriber 200 communicates to the IVR 180, the IVR 180 may determine that the Subscriber 200 is calling from Pakistan. This information may be used to specify the consumer's location to the Proprietary HTTP Server 125. Based on this information, the Service Provider HTTP Server 120 may be able to utilize corresponding proprietary features on Vocabulary Domain Based Text-to-Speech Converter 210 to enable the IVR 180 to retrieve the audio file, which may contain the latest news stories for Pakistan from the Proprietary HTTP server 125.

[0038] Reference is now made to Fig. 3, which depicts an efficient mechanism for providing vocabulary domain text-to-speech services. A Client 300 preferably sends textual information to the TTS Server 220 to be processed. A Parser 310, located within the TTS server 220, preferably receives the textual information and parses the text into phrases. A Text Distributor 320, also located within the TTS server 220, preferably first checks with a Cache 330, located within the TTS server 220, to determine whether the phrases have been previously cached. If so, the Cache 330 may return the audio content back to the Client 300. Otherwise, the Text Distributor 320 may map phrases to their

respective clusters, which may have been previously defined by the Analyzer/Vocabulary Domain Definer 150 (Figs. 1 and 2).

[0039] Each cluster may be associated with a representative Limited Vocabulary Domain Server 340. The Text Distributor 320 may enqueue the phrases on one of a plurality of Queues 350, each associated with the respective limited vocabulary domain. Each Queue 350 may have associated therewith a Thread Pool 360 and a TTS Client 370 to facilitate distributed concurrent processing of requests.

[0040] When the Text Distributor 320 enqueues a phrase on a particular Queue 350, the relevant Queue 350 may notify the Thread Pool 360 of the new phrase. Should the Thread Pool 360 have a free thread, the Thread Pool 360 may dequeue the phrase from the Queue 350 and may communicate the phrase to the TTS Client 370. The TTS Client 370 may further transmit the phrase to the relevant Limited Vocabulary Domain Server 340. The Limited Vocabulary Domain Server 340 is preferably defined to have a limited vocabulary domain and to be capable of suitably processing the phrase and converting the phrase to audio content. The phrase may be stored in the Cache 330 for future reference and may be transmitted back to the Client 300.

[0041] It will be appreciated by persons skilled in the art that the present invention is not limited by what has been particularly shown and described hereinabove. Rather the present invention includes combinations and sub-combinations of the various features described hereinabove as well as modifications and extensions thereof, which would occur to a person skilled in the art and which do not fall within the prior art.